Exhibit 4

Harvard Business School

Alain Bonacossa
he/him/his
Research Integrity Officer,
Division of Research and Faculty Development

617.496.6348 (o)
917.292.8375 (m)
abonacossa@hbs.edu

CONFIDENTIAL

Date: June 13, 2023

By Email:
Dr. Patricia J. Bauer, patricia.bauer@emory.edu

Dear Dr. Bauer,

I am writing to inform you that Harvard Business School (HBS) has reviewed concerns about certain data previously published by Dr. Francesca Gino in the following articles:

> Gino, F., Kouchaki, M., & Galinsky, A. D. (2015). The moral virtue of authenticity: How inauthenticity produces feelings of immorality and impurity. *Psychological Science*, 26(7), 983–996
>
> Gino, F., & Wiltermuth, S. S. (2014). Evil genius? How dishonesty can lead to greater creativity. *Psychological Science*, 25(4), 973–981

We have included 2 appendices to this letter. The first appendix relates to the 2015 publication listed above includes (a) a description of the data anomalies for Study 4 observed in the Open Science Framework (OSF) dataset, (b) a comparison of the original data for Study 4 gathered using Qualtrics with the data observed in OSF, and (c) an assessment by an independent forensic firm. In summary, a comparison of the original Qualtrics dataset with the dataset posted on OSF revealed discrepancies that strongly support the hypothesized and reported results. Additionally, when the analyses reported in the published paper were run on the original Qualtrics data, the key result – that participants in the pro-attitudinal condition expressed significantly lower desirability of cleaning products – failed to replicate.

The second appendix relates to the 2014 publication listed above and includes (a) a description of observed data anomalies for Study 4, (b) a comparison of the earliest known version of data for Study 4 with the last known dataset used for publication, and (c) an assessment by an independent forensic firm. In summary, the comparison of the earliest known data with the dataset used for publication revealed three anomalies in the earliest versions of the data available which impacted the reported results in the direction of the hypothesized results.

We thus believe the results reported in Study 4 of the 2015 article and in Study 4 of the 2014 article are invalid due to alterations of the data that affects the significance of the findings. We are informing the co-authors of both articles, and are recommending that both articles be retracted. If you wish to discuss this matter further or if you have any questions, please feel free to reach out to me at 617-496-6348 or abonacossa@hbs.edu.

Sincerely,

Alain Bonacossa
Research Integrity Officer

APPENDIX REGARDING
Gino, F., Kouchaki, M., & Galinsky, A. D. (2015). The moral virtue of authenticity: How inauthenticity produces feelings of immorality and impurity. Psychological Science, 26(7), 983–996.

Data anomalies observed in the OSF dataset (Study 4)

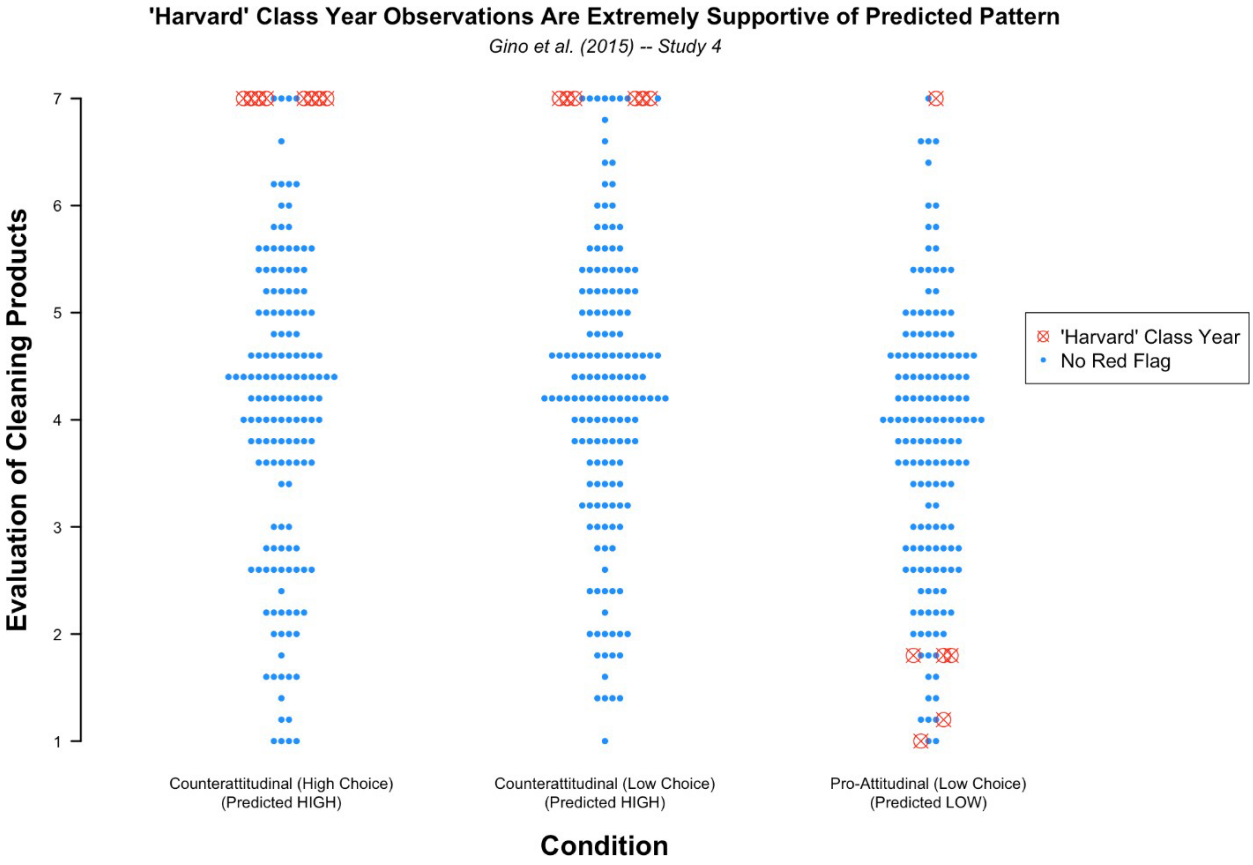
Data were retrieved from the OSF (https://osf.io/sd76g), where they have been posted since 2015. The anomaly in this dataset involves how some students answered Question #6: "Year in School."

The screenshot below shows a portion of the dataset. In the "yearSchool" column, one can see that students approach this "Year in School" question in a number of different ways. For example, a junior might write "junior", or "2016" or "class of 2016" or "3" (to signify that they are in their third year). All of these responses are reasonable.

A less reasonable response is "Harvard", an incorrect answer to the question. One would not expect many students to independently make the same idiosyncratic mistake, but the data file indicates that 20 students did so. Those students' responses are very close to one another, all within 35 rows (450 through 484) in the posted dataset:

1	instr	college_st	inFavor_in	strongOp	age	male	gender_T	yearSchool	condition
443	1	1	1	7	19	1		Sophomore	ProAttitudinal
444	1	1	1	7	20	1		Junior	No_Choice
445	1	1	1	6	19	0		sophomore	High_Choice
446	1	1	1	6	20	1		Junior	ProAttitudinal
447	1	1	1	7	21	1		Senior (Class of 201	No_Choice
448	1	1	1	5	22	1		Senior	High_Choice
449	1	1	1	5	21	1		Senior	ProAttitudinal
450	1	1	1	7	23	0		harvard	No_Choice
451	1	1	1	4	21	0		2015	High_Choice
452	1	1	1	7	20	1		Junior	No_Choice
453	1	1	1	7	18	0		Sophomore	ProAttitudinal
454	1	1	0	7	25	0		Harvard	High_Choice
455	1	1	0	7	25	0		Harvard	ProAttitudinal
456	1	1	1	7	22	1		Harvard	ProAttitudinal
457	1	1	0	7	24	0		Harvard	High_Choice
458	1	1	1	7	22	0		Harvard	High_Choice
459	1	1	0	7	25	0		Harvard	No_Choice
460	1	1	1	7	23	1		Harvard	ProAttitudinal
461	1	1	0	7	25	0		Harvard	High_Choice
462	1	1	0	6	25	1		4	No_Choice
463	1	1	0	7	24	0		Harvard	No_Choice
464	1	1	1	5	23	0		1	High_Choice
465	1	1	1	4	19	0		Sophomore	No_Choice
466	1	1	1	6	28	1		5	High_Choice
467	1	1	1	6	22	1		Senior	ProAttitudinal
468	1	1	1	6	20	0		Junior	High_Choice
469	1	1	1	5	23	1		2015	High_Choice
470	1	1	1	6	22	1		Senior	No_Choice
471	1	1	1	6	22	0		2015/Senior	ProAttitudinal
472	1	1	1	6	36	1		2010	High_Choice
473	1	1	1	7	25	0		Harvard	ProAttitudinal
474	1	1	0	5	25	0		Harvard	High_Choice
475	1	1	1	7	22	1		Harvard	No_Choice
476	1	1	1	7	23	1		Harvard	High_Choice
477	1	1	0	7	25	0		Harvard	ProAttitudinal
478	1	1	0	7	26	1		Harvard	No_Choice
479	1	1	1	6	20	0		2013	No_Choice
480	1	1	0	6	21	0		2012	ProAttitudinal
481	1	1	1	7	24	1		Harvard	High_Choice
482	1	1	1	7	27	0		Harvard	ProAttitudinal
483	1	1	1	7	25	1		Harvard	High_Choice
484	1	1	1	7	27	0		Harvard	No_Choice
485	1	1	1	7	26	1		4	High_Choice
486	1	1	0	6	22	0		2012	High_Choice
487	1	1	1	6	20	1		2013	No_Choice

If these peculiar observations were tampered with, then one might see that students who answered "Harvard" were especially likely to confirm the authors' hypothesis. The figure below presents a Bee Swarm plot, which depicts each observation in the dataset, separately for each experimental condition. The plot depicts the key dependent variable, participants' average ratings of how much they desired five cleaning products. The 20 "Harvard" observations are represented as red X's, while the rest of the observations are represented as blue dots:



Here one can see that in the two counter-attitudinal conditions, which were predicted to induce a desire for cleaning products and thus higher values on y-axis, every "Harvard" observation has the highest possible average value (i.e., a 7.0). Conversely, in the pro-attitudinal condition, which was predicted to induce a lower desire for cleaning products, every "Harvard" observation is associated with a low value, except for one (which itself happens to be the only one associated with a lowercase "harvard").

The difference between the pro-attitudinal and counter-attitudinal conditions for just these 20 observations is highly significant, with a p-value indicating that it would occur by chance less than one in a million times: $t(18) = 7.84$, $p < .000001$.

Comparison of observations between the original Qualtrics dataset and OSF dataset (Study 4)

Among the observations that list "Harvard" as their answer to the "Year in School" question, none had a "college.harvard.edu" email address. In contrast, most of the observations that did not answer "Harvard" as the "Year in School" provided a Harvard email address (e.g., one ending in "college.harvard.edu"). The responses by the "Harvard group" on the key dependent variable ─ average ratings of desire for the five cleaning products ─ were of highly similar magnitudes and influenced the overall experimental findings in the hypothesized direction.

A discrepancy was observed in the N for the dataset obtained from Dr. Gino's research records and the N for the publicly posted dataset available on OSF, which was analyzed previously. Dr. Gino's file (from her research records) showed 455 responses to the information requests (e.g., age, gender, year in school) and the experiment's questions. The OSF dataset had 491 responses.

In a direct comparison of the two data sets (the file from the research records and the OSF file), the following were observed:

1. Some participants in Dr. Gino's file were not in the OSF file. For example, Dr. Gino's dataset included 24 participants who responded with "Harvard" to the "Year in School" question while the publicly available dataset on OSF included only 20 participants who responded with "Harvard" as their year in school. These were not the only instances of participants in Dr. Gino's file that did not appear in the OSF file.
2. Some participants in the OSF file were not in Dr. Gino's file.

In addition, other anomalies were observed in both data sets, especially from people that did not report a college.harvard.edu email address in Dr. Gino's file. For example, several otherwise identical records in both data sets differ only in the scores reported about a participant's preference for "clean products."

Assessment by an Independent Forensic Firm

Executive Summary.

There appears to be multiple study approaches employed asynchronously following initial participant recruitment and assessment which produced data sets that are related to the data published in Gino, Kouchaki, & Galinskyn (2015), "The moral virtue of authenticity: How inauthenticity produces feelings of immorality and impurity," *Psychological Science*, 26(7), 983–996 (the **2015 PS paper)**. However, while the data sets (**LAB** and **ONLINE data**) have some shared features, the data analyzed show other inconsistencies. Specifically:

1. Despite the fact that the two data sets are derived from distinct protocols, an amalgamation of these two resultant LAB and ONLINE data sets appear to be the source of a third data set (the **OSF data** set) published with the 2015 PS paper (see **section III.** below).
2. The number of participants published is lower than what appear to be acquired in the available Qualtrics surveys (see **Observations 2**).
3. Analysis of the datasets demonstrate:
    a. there were data points found in the **OSF data** set that were not found in either of these apparent source data files, the **LAB** and **ONLINE data**, (see Figure 1b below for an example, **Observations 3**),
    b. when the published **OSF data** (.sav) were exported in Excel and re-evaluated with the authors protocols, there is an example of resultant calculated data (average) published that do not appear to align with the source data that was calculated in the same way (e.g., published average ≠ source data average, see **Observations 1.** below for the results). It is not clear how this alteration was introduced into the data set,
    c. not all **LAB** and **ONLINE data** were sourced to create the **OSF data** set, and there were no clear inclusion and/or exclusion criteria applied for utilizing these data sources (see example Figure 1a. below), and
    d. when a common data set was extrapolated from the **LAB** and **ONLINE data** (e.g., "**Qualtrics data**", n=530) using apparently completed survey data available from participants and compared to **OSF data** (.sav file, n=491) using the authors protocols (.sps file found in OSF), the statistical outcomes do not appear to align with what are reported. Additionally, the statistical outcomes, as derived from the "Qualtrics data" source data, appear to affect at least one of the author's published conclusions stated in the 2015 PS paper (see example Table 2. below, in **Observations 4**).

ANALYSIS AND OBSERVATIONS

Data Analysis.

Data from three sources were compared: publicly available Experiment 4 (**OSF data**) in comparison to reported original Qualtrics data sources (**ONLINE data** and **LAB data**).

Approach:
The **LAB data** file included 69 input columns and the **ONLINE data** file included 109 input columns.

25 of the input columns in both the **LAB data** and **ONLINE data** files were found in the **OSF data.** The **OSF data** also included 5 analysis columns that were not present in the **LAB** or **ONLINE data** files.

Three distinct sets of comparisons/analysis were performed:

i. Re-calculation of the analysis columns included in the **OSF dataset**
ii. Comparison of the **ONLINE** and **LAB data** with the **OSF** data to identify the full set of survey responses that appeared to align and potential survey responses that were altered, added, and/or removed.
iii. Statistical analysis of the **ONLINE** and **LAB** data, and **OSF data,** to identify aligning and non-aligning data entries.

Replication

The **OSF data** file contained 5 analysis columns . These columns were recalculated to compare them with the results reported online. For consistency with the methodology utilized by the authors, the calculations were adapted from the authors' own protocol file.

Comparison of OSF and Qualtrics files

Since the **Qualtrics data** (LAB and ONLINE) contained higher number of input columns, data were compared within the shared subset of input columns between the **Qualtrics data** and the **OSF data** to determine which subset of values, if any, were utilized to compile the **OSF data**.

Observations.

Identifying Differences between Qualtrics data and OSF data

The available datasets were subdivided into subgroups:

1- The survey scores from **ONLINE data** that matched the **OSF data** (N = 391)
2- The survey scores from **LAB data** that matched the **OSF data** (N = 75)
3- The survey scores from the **OSF data** with no exact match in Qualtrics (N = 25)

4- The survey scores from **ONLINE data** that had no match in OSF ()
5- The survey scores from **LAB data** that had no match in OSF (N = 3)

Figure 1 shows a score data plot that outlines the data apparently absent in the **OSF data** that were present in the identified source data (**LAB** and **ONLINE data** sets.) Furthermore, there are data present in the **OSF data** that are not found in either the **LAB** or **ONLINE data**. As the different Qualtrics datasets were identified as the source data it is not clear where the additional data present in the OSF data derive from, and the rationale (or description) of the excluded data points from the source data (**LAB** and **ONLINE data**) are equally unknown.

If one were to define "lower scale" data as ≤ 3.5 on scale of 1-7 and "higher scale" data as > 3.5, as shown in the Figure1a, it would appear that as the data transitioned from "source" to "published" data (see blue and yellow data points) higher scale response data available in "source" data were absent in the "published" data. While in Figure 1b ("published"), there are data that don't seem to derive from the available research record (**GREEN** data points, n=25 total data points were identified):

- higher scale response data were added to conditions 0 and 1, and
- lower scale response data were added for condition 2

In some cases, there appears to be some trends in data modifications which may align with anticipated outcomes, the data apparently added into, and removed from, the published data in alignment with the anticipated outcome for the condition (lower desire for cleanliness in the pro attitudinal group of Condition 2). For example, for Condition 2, nine "lower scale" data points [≤ 3.5 on scale of 1-7] were added and additional "higher scale" data [> 3.5] were removed. While these modifications do not seem to have an impact on the directionality of the data with respect to the authors' anticipated outcomes, and this is across all categories and all conditions, the modifications of the data appear to have an impact in at least one area of the authors' outcomes discussed in the **2015 PS paper**.

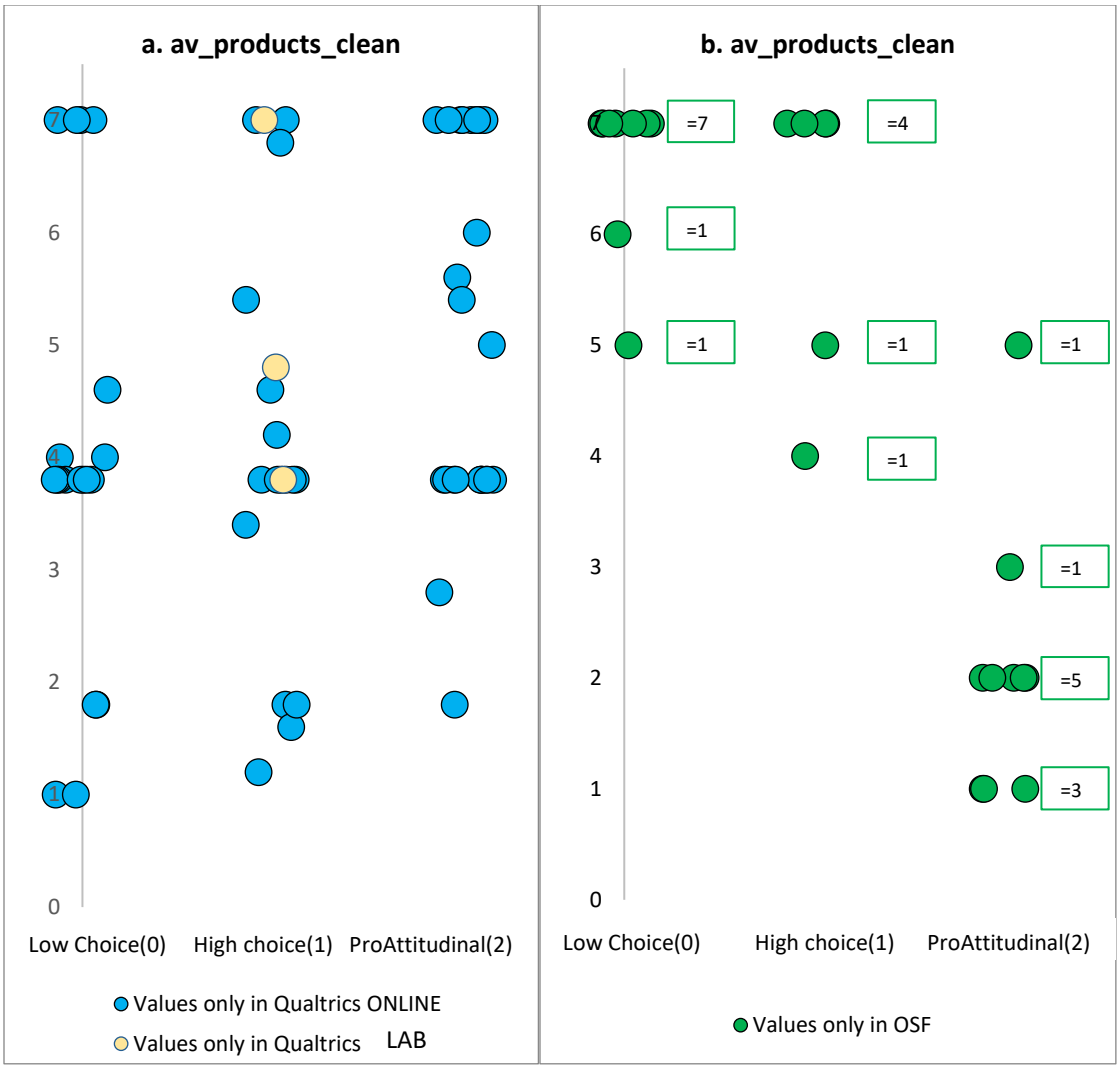


Figure 1. "cleaning products". **a.** Visualization of the trend of data found only in LAB (YELLOW) or ONLINE (**BLUE**) data and absent from the published OSF data. Applying a randomization factor around the axis of the specific condition to spread the values on the x axis and allow for better data visualization of the scale points added (or absent) per data set. **b.** Data found only in the OSF data and absent in the LAB and/or ONLINE data in **GREEN** (source unknown) are demonstrated to evaluate possible trends per Condition.

An additional check was conducted to evaluate if the irregularities were to be attributed to surveys having 'harvard' (or 'Harvard') as an entry on the 'yearSchool' column. A search through the **ONLINE data** revealed 24 instances where participants apparently wrote 'harvard' as school year. The 24 participants who wrote 'harvard' as school year in the *ONLINE data* did not correspond exactly with those in the *OSF data* (e.g., 8 of the OSF entries used in the paper ≠ ONLINE entries for 'harvard'). However, the rationale for inclusion or exclusion of value data in the **OSF data** set did not appear to be reliant on this category.

Assessing Differences between Qualtrics and OSF data

According to p.992 of the 2015 PS Paper:

*"**Manipulation check: self-alienation.** A one-way ANOVA using self-alienation as the dependent measure revealed a main effect of condition, $F(2, 487) = 21.14$, $p < .001$, $\eta_p^2 = .08$. Pairwise comparisons (with Bonferroni adjustment) revealed that participants reported lower self- alienation in the proattitudinal condition ($M = 1.88$, $SD = 0.87$) than in both the high-choice, counterattitudinal condition ($M = 2.56$, $SD = 1.31$; $p < .001$) and the low-choice, counterattitudinal condition ($M = 2.70$, $SD = 1.40$; $p < .001$). Participants reported the same perceived self-alienation in the two counterattitudinal conditions ($p = .94$).*

***Perceived choice.** A one-way ANOVA using perceived amount of choice as the dependent measure revealed a main effect of condition, $F(2, 487) = 62.35$, $p < .001$, $\eta_p^2 = .20$. Pairwise comparisons (with Bonferroni adjustment) revealed that participants reported lower perceived choice in the low-choice, counterattitudinal condition ($M = 2.85$, $SD = 1.98$) than in the high-choice, counterattitudinal condition ($M = 3.63$, $SD = 2.16$; $p = .001$) and in the proattitudinal condition ($M = 5.24$, $SD = 1.78$; $p < .001$). Perceived choice was higher in the proattitudinal condition than it was in the high-choice, counterattitudinal condition ($p < .001$).*

***Desirability of cleansing products.** A one-way ANOVA using participants' desirability ratings of cleansing products as the dependent measure revealed a main effect of condition, $F(2, 487) = 8.24$, $p < .001$, $\eta_p^2 = .033$. Pairwise comparisons (with Bonferroni adjustment) revealed that participants reported less desire for cleansing products in the proattitudinal condition ($M = 3.72$, $SD = 1.33$) than in both the high-choice, counterattitudinal condition ($M = 4.18$, $SD = 1.51$; $p = .012$) and the low-choice, counterattitudinal condition ($M = 4.34$, $SD = 1.44$; $p < .001$). Desirability ratings of cleansing products did not differ between the latter two conditions ($p = .94$). There were no differences across conditions in desirability ratings of the non-cleansing products, $F(2, 487) = 1.21$, $p = .30$, $\eta_p^2 = .005$."*

The replication of the statistical assessment of the data relative to Experiment 4 shows lower statistical significance between samples when comparing results obtained under the three conditions, however, general statistical trends appear to be retained for 3 of the 4 categories.

Figure 2. Desirability of cleaning products

	Average and Standard Deviation						p between conditions (Bonferroni)		
	ProAttitudinal		high choice		low-choice		proAttitudinal vs lowChoice	proAttitudinal vs highChoice	low-high
	M	SD	M	SD	M	SD			
Results reported (N=491)	3.72	1.33	4.18	1.51	4.34	1.44	<0.001	0.012	0.94
Results obtained OSF (N=490)	3.72 (N=161)	1.33	4.18 (N=161)	1.51	4.34 (N=168)	1.44	0.0003003	0.012	0.94
Combined Qualtrics sets (N=529)	4.03 (N=176)	1.40	4.11 (N=175)	1.51	4.19 (N=178)	1.42	0.8661559	1.000	1.00

However, as seen in Figure 2, the modifications of the data set appear to have an impact on the authors' assessments of the experimental outcomes for "desired cleanliness". During the discussion the authors conclude that "When participants wrote essays that were not consistent with their internal beliefs, regardless of choice, they showed a greater desire for cleanliness." However, when comparing the OSF to the Qualtrics data source(s), there does not appear to be any significantly greater desire for cleanliness regardless of the essay type either (e.g., see proAttitudinal vs lowChoice and proAttitudinal vs highChoice conditional comparisons, p=0.87 and p=1, respectively).

"Harvard" data

As described, the data included (or not) in this category did not appear to align across the data sets; 12 of the participants who wrote 'harvard' as school year in the **ONLINE data** sheet appeared to correspond with participants included in the **OSF data** sheet (3 for condition 0, 8 for condition 1, 1 for condition 2). However, as reviewed, the rationale for inclusion or exclusion of value data in the OSF data did not appear to be reliant on this category (or any specific category or grouping). Additionally, inclusion (or exclusion) of "Harvard" entries does not appear to alter the age of included participants in the **OSF data** set even if in some cases there is an impact on the statistical significance. However, given that these data already appear to represent a data set that do not align with source files and the research record (e.g., OSF vs Qualtrics Data), it is difficult to determine what impact (if any) participant data associated with this category have on the overall study and its reported outcomes.

Summary.

There appears to be inconsistencies within the available data sets related to the data published in the 2015 PS paper. Multiple study approaches, modified in progress, resulted

in the production of two different data sets (e.g., as related to LAB and ONLINE data sets) which may confound the assessments published in the 2015 PS paper. Furthermore, there appears to be additional value modifications as well as addition and removal of data within these data sets that have an impact on the authors' conclusionary statements regarding Experiment 4. The rationale for the modifications within the data as they transition from the apparent research record into the published record were less clear (e.g., the modifications, while having some directionality, did not appear to align completely with authors hypothesized outcomes). However, the modifications did have an impact on at least one aspect of the authors' conclusions of the study; there do not appear to be any statistically significant differences regarding the desire for cleanliness associated with choice and internal beliefs based on the calculations of the available research record.

APPENDIX REGARDING
Gino, F., & Wiltermuth, S. S. (2014). Evil genius? How dishonesty can lead to greater creativity. Psychological Science, 25(4), 973–981.

Data anomalies observed in the dataset (Study 4)

Direct Evidence of Tampering

The dataset seems to be sorted by two columns, first by a column called "cheated", indicating whether participants cheated on the coin toss task (0 = did not cheat; 1 = cheated), and then by a column called "Numberofresponses", indicating how many uses for a newspaper the participant generated.

For example, the screenshot below depicts the first 40 observations in the dataset. Because the data are sorted first by the "cheated" column, all of these observations represent non-cheaters (i.e., scores of 0 in that "cheated" column). The shown rows are perfectly sorted by the "Numberofresponses" column. All of the 135 non-cheaters in the dataset are sorted by the "Numberofresponses" column.

	StartDate	EndDate	MTurkID	Cum_ID	cheated	Numberofresponses
2	11/17/12 23:54	11/18/12 0:07	AD8VVYGP4LRKG	144	0	2
3	11/17/12 23:17	11/17/12 23:41	A2KJZAMH6G8LWC	91	0	2
4	11/17/12 23:44	11/17/12 23:57	A21TECY6SM7BNV	127	0	3
5	11/17/12 22:57	11/17/12 23:11	A2GRSJHXTR7JQR	24	0	3
6	11/18/12 0:00	11/18/12 0:20	A1FAQI6Q4WCS	168	0	3
7	11/17/12 23:41	11/17/12 23:52	A1YZJ7OO7Q2D89	113	0	3
8	11/17/12 23:37	11/17/12 23:47	AVA93G56VQLZA	101	0	3
9	11/17/12 23:20	11/17/12 23:32	A20863XUQJT5T1	76	0	3
10	11/18/12 0:11	11/18/12 0:24	A27I79P03I0ZPO	173	0	3
11	11/17/12 23:11	11/17/12 23:28	A12WY0ZDGVOZQS	69	0	3
12	11/17/12 23:41	11/17/12 23:56	A20552JTR91G67	124	0	3
13	11/17/12 23:17	11/17/12 23:33	A3Q9UUF8RPV4LQ	79	0	3
14	11/17/12 22:49	11/17/12 22:58	A2BH9W7Y1TL3X8	1	0	3
15	11/17/12 23:59	11/18/12 0:10	A034420738QHAX9TNO9BA	152	0	4
16	11/17/12 23:38	11/17/12 23:51	a32k7qy8nwzx43	110	0	4
17	11/17/12 23:05	11/17/12 23:23	A2DAT0DBUXU8FF	55	0	4
18	11/17/12 23:39	11/17/12 23:49	A20A0EM29IJLSK	103	0	4
19	11/17/12 23:31	11/17/12 23:51	APJEYYRENCAC6	109	0	4
20	11/17/12 23:02	11/17/12 23:27	A1L6EDKEUG69XB	66	0	4
21	11/18/12 0:00	11/18/12 0:10	AYZ00GXI5D15Y	150	0	4
22	11/17/12 23:22	11/17/12 23:35	APHNYDGTCRN3O	82	0	4
23	11/17/12 23:19	11/17/12 23:32	A1MM8TSLCHVMNK	75	0	4
24	11/17/12 23:12	11/17/12 23:24	A3AZJGI9D7C0PD	57	0	4
25	11/17/12 22:52	11/17/12 23:17	A3DQUF5TM9VTS7	37	0	4
26	11/17/12 23:50	11/18/12 0:03	A77M840AXJ16B	137	0	4
27	11/18/12 0:02	11/18/12 0:10	A3GSCVUHX7DM8T	151	0	4
28	11/17/12 23:05	11/17/12 23:24	A26L91YL0GDGD8	58	0	4
29	11/17/12 23:27	11/17/12 23:53	AJY9CIX7FW9W1	115	0	4
30	11/17/12 23:48	11/18/12 0:02	ALSE4C4Q3R6G	133	0	5
31	11/17/12 22:54	11/17/12 23:08	A2R8SVW42IYFYX	17	0	5
32	11/17/12 22:59	11/17/12 23:17	A07109741WN0LPDUN9GL9	34	0	5
33	11/17/12 23:25	11/17/12 23:37	A1GFD4B3NOMWIY	86	0	5
34	11/17/12 23:37	11/17/12 23:54	ADQML8ECWYME5	119	0	5
35	11/17/12 23:04	11/17/12 23:32	A3FAAKASDY5HE6	183	0	5
36	11/17/12 22:55	11/17/12 23:14	A5SUR5C68YYN8	30	0	5
37	11/17/12 22:56	11/17/12 23:08	A2MBAN2GDK1P1J	16	0	5
38	11/17/12 23:48	11/18/12 0:00	A34N9G0IEI28IG	131	0	5
39	11/17/12 23:46	11/18/12 0:06	A3AHNUDEOZ33JE	143	0	5
40	11/17/12 23:25	11/17/12 23:38	A1QK6O24KDVLJ1	88	0	5
41	11/17/12 22:58	11/17/12 23:19	A7NLUN5YH4S9L	43	0	5

The next screenshot shows that while 43 cheaters are also sorted by this variable, there are 13 observations (highlighted in yellow) that are not in the same sort order.

1	StartDate	EndDate	MTurkID	Cum_ID	cheated	Numberofresponses
132	11/18/12 0:04	11/18/12 0:13	A1X82CGYFM586F	155	0	11
133	11/17/12 23:08	11/17/12 23:22	A1F14BB4PV053A	53	0	11
134	11/17/12 23:22	11/17/12 23:37	A356ZZWYC8GRVY	85	0	11
135	11/17/12 23:44	11/18/12 1:05	A34DG3IZ88WWBT	192	0	12
136	11/17/12 22:58	11/17/12 23:14	A3P7XKTEBOKN5R	29	0	13
137	11/18/12 0:01	11/18/12 0:20	ADTNOFJHTTB1L	167	1	3
138	11/17/12 23:34	11/17/12 23:53	A1UNAJF3E5HH17	114	1	3
139	11/17/12 23:44	11/17/12 23:57	A0377367199XXE56OT9GZ	126	1	4
140	11/17/12 23:36	11/17/12 23:46	A2DUKWR9I6FFZV	99	1	4
141	11/17/12 23:02	11/17/12 23:17	AE3D6SE2D8UPQ	36	1	13
142	11/17/12 23:32	11/17/12 23:43	A21MCWTDIKATV5	97	1	9
143	11/17/12 23:59	11/18/12 0:10	A28XLOE0FMG1ZX	153	1	5
144	11/17/12 22:55	11/17/12 23:04	A126XP3VIWJKD6	8	1	5
145	11/18/12 0:07	11/18/12 0:21	A3E1EPRY1OYE34	171	1	9
146	11/17/12 23:30	11/18/12 0:03	A27AEIRFEFR4US	136	1	5
147	11/17/12 23:30	11/18/12 0:44	A07854333QXCSICF01THG	191	1	9
148	11/17/12 23:38	11/17/12 23:50	A311BZDLCK6HQQ	105	1	8
149	11/17/12 22:59	11/17/12 23:15	A1ILA0RGDB9JJ6	32	1	9
150	11/17/12 23:11	11/17/12 23:22	A22LZ62E0UC4VL	51	1	5
151	11/17/12 23:49	11/18/12 0:03	A1SHH0U3JHSCSV	187	1	6
152	11/18/12 0:03	11/18/12 0:22	A37JD0XUZHQYRC	172	1	6
153	11/17/12 22:52	11/17/12 23:04	ALML8V38FDV0	9	1	9
154	11/17/12 23:58	11/18/12 0:14	A3W4736CCV8TT4	157	1	11
155	11/18/12 0:07	11/18/12 0:15	AUN8AE8UCO3MD	159	1	14
156	11/17/12 23:13	11/17/12 23:29	Jazzy67033	180	1	6
157	11/17/12 22:58	11/17/12 23:08	A208MTGA7V29TP	14	1	8
158	11/17/12 23:51	11/18/12 0:08	A2UL07RCD2RO8R	146	1	10
159	11/17/12 22:51	11/17/12 23:10	AP37A6DG5TTEM	20	1	7
160	11/18/12 0:03	11/18/12 0:14	A2H18EYM79ZRCW	156	1	7
161	11/17/12 23:59	11/18/12 0:09	A1BCCFEEN32OWP	149	1	8
162	11/17/12 23:03	11/17/12 23:15	A3TN3GQAO61BVB	31	1	7
163	11/18/12 0:03	11/18/12 0:21	hhendric@hotmail.com	169	1	7
164	11/17/12 23:13	11/17/12 23:26	A62RZY5BWOZZM	63	1	14
165	11/17/12 23:25	11/17/12 23:47	AVUAN8WKJ443M	102	1	8
166	11/17/12 23:48	11/17/12 23:59	A25KU26Y8FTJPV	129	1	8
167	11/17/12 22:55	11/17/12 23:06	A3OF0DCN3KU8HT	11	1	8
168	11/17/12 23:11	11/17/12 23:18	A47QHTQNUTOVL	42	1	8
169	11/17/12 23:52	11/18/12 0:03	A1ASPIEIOZXL3U	138	1	8
170	11/17/12 23:27	11/17/12 23:32	A3E0AY1XXP8IBQ	77	1	9
171	11/17/12 23:57	11/18/12 0:21	A1R7CJMWXC79UO	170	1	10
172	11/17/12 23:03	11/17/12 23:10	A5VWAZZ49D5WU	22	1	10
173	11/17/12 23:21	11/17/12 23:31	AGX6FRHVVU2WS	74	1	10
174	11/17/12 23:25	11/17/12 23:37	A24JC2CF7MMG41	84	1	10
175	11/17/12 23:46	11/17/12 23:58	A1REWUVT3N8SN7	128	1	11
176	11/17/12 23:37	11/17/12 23:50	A27MJOV91GA8R3	106	1	11
177	11/17/12 23:06	11/17/12 23:17	A17M7G85OEIB3U	35	1	11
178	11/17/12 23:06	11/17/12 23:21	A2IF1VIC7GZUN	50	1	12
179	11/17/12 23:07	11/17/12 23:17	A2GPIQQ2PJ87Q0	38	1	13

It is not possible to sort the dataset to generate the order in which the data were saved. They were either originally entered this way (which does not seem plausible, since the data originate in a Qualtrics file, which by default sorts by time), or they were manually altered.

Second, because rows are sorted by the variable of interest, "numberOfUses", if the values that are out of order were changed, it is straightforward to impute what they were changed from. For example, row #141 is "13", the number right before it is "4", and the first value in the expected sort order after it is "5". Therefore, if the data were changed, then we can assume that that "13" used to be either a "4" or a "5".

One can do this for each of the 13 highlighted values in the dataset, and thus reconstruct what the data

looked like before they were altered. The screenshot below shows the imputed values for all relevant cells. The first new column ("Imputed1") imputes the lowest value that is consistent with the neighboring observations, and the second new column ("Imputed2") shows the highest value. So we see, for example, that that first "13" could have been either a "4" or a "5".

1	StartDate	EndDate	MTurkID	Cum_ID	cheated	Numberofresponses	Imputed1	Imputed2
137	11/18/12 0:01	11/18/12 0:20	ADTNOFJHTTB1L	167	1	3	3	3
138	11/17/12 23:34	11/17/12 23:53	A1UNAJF3E5HH17	114	1	3	3	3
139	11/17/12 23:44	11/17/12 23:57	A0377367199XXE56OT9GZ	126	1	4	4	4
140	11/17/12 23:36	11/17/12 23:46	A2DUKWR9I6FFZV	99	1	4	4	4
141	11/17/12 23:02	11/17/12 23:17	AE3D6SE2D8UPQ	36	1	13	4	5
142	11/17/12 23:32	11/17/12 23:43	A21MCWTDIKATV5	97	1	9	4	5
143	11/17/12 23:59	11/18/12 0:10	A28XLOE0FMG1ZX	153	1	5	5	5
144	11/17/12 22:55	11/17/12 23:04	A126XP3VIWJKD6	8	1	5	5	5
145	11/18/12 0:07	11/18/12 0:21	A3E1EPRY1OYE34	171	1	9	5	5
146	11/17/12 23:30	11/18/12 0:03	A27AEIRFEFR4US	136	1	5	5	5
147	11/17/12 23:30	11/18/12 0:44	A07854333QXCSICF01THG	191	1	9	5	5
148	11/17/12 23:38	11/17/12 23:50	A311BZDLCK6HQQ	105	1	8	5	5
149	11/17/12 22:59	11/17/12 23:15	A1ILA0RGDB9JJ6	32	1	9	5	5
150	11/17/12 23:11	11/17/12 23:22	A22LZ62E0UC4VL	51	1	5	5	5
151	11/17/12 23:49	11/18/12 0:03	A1SHH0U3JH5CSV	187	1	6	6	6
152	11/18/12 0:03	11/18/12 0:22	A37JD0XUZHQYRC	172	1	6	6	6
153	11/17/12 22:52	11/17/12 23:04	ALML8V38FDV0	9	1	9	6	6
154	11/17/12 23:58	11/18/12 0:14	A3W4736CCV8TT4	157	1	11	6	6
155	11/18/12 0:07	11/18/12 0:15	AUN8AE8UCO3MD	159	1	14	6	6
156	11/17/12 23:13	11/17/12 23:29	Jazzy67033	180	1	6	6	6
157	11/17/12 22:58	11/17/12 23:08	A208MTGA7V29TP	14	1	8	6	7
158	11/17/12 23:51	11/18/12 0:08	A2UL07RCD2RO8R	146	1	10	6	7
159	11/17/12 22:51	11/17/12 23:10	AP37A6DG5TTEM	20	1	7	7	7
160	11/18/12 0:03	11/18/12 0:14	A2H18EYM79ZRCW	156	1	7	7	7
161	11/17/12 23:59	11/18/12 0:09	A1BCCFEEN32OWP	149	1	8	7	7
162	11/17/12 23:03	11/17/12 23:15	A3TN3GQAO61BVB	31	1	7	7	7
163	11/18/12 0:03	11/18/12 0:21	hhendric@hotmail.com	169	1	7	7	7
164	11/17/12 23:13	11/17/12 23:26	A62RZY5BWOZZM	63	1	14	7	8
165	11/17/12 23:25	11/17/12 23:47	AVUAN8WKJ443M	102	1	8	8	8
166	11/17/12 23:48	11/17/12 23:59	A25KU26Y8FTJPV	129	1	8	8	8
167	11/17/12 22:55	11/17/12 23:06	A3OF0DCN3KU8HT	11	1	8	8	8

When one reconstructs the data in this way, by replacing the highlighted values with the values one would impute based on the order in which data are sorted, the significant relationship between cheating and creativity on the uses task entirely disappears. Its p-value goes from <.0001 to .292 ("Imputed1") or .180 ("Imputed2").

Comparison of observations between the earliest known dataset and the last known dataset on which publication was based (Study 4)

The mean "# Responses" score of "in-sequence" observations was 7.5, while the mean "# Responses" score of "out-of-sequence" observations was much higher, at 10.1. Replacing an out-of-sequence entry in the "# responses" column with an adjacent "in sequence" score caused the mean score of respondents in the Cheating condition to decrease from 8.3 to 7.0, greatly closing the gap to the mean score of 6.5 for Honest respondents.

The data file from Dr. Gino's research records, which she identified as the file for this experiment, contains data for 178 participants. However, the published paper reports 208 participants.

Assessment by an Independent Forensic Firm

Executive Summary.

The analysis of files demonstrated an apparent series of manipulations to a dataset prior to its publication as Experiment 4 in the 2014 *Psychological Science* paper:

- In earlier versions of source documentation available at the time of this report, a series of data, which were already color coded by an unknown individual, demonstrate manual alterations of data points that ultimately appear in the dataset that Dr. Gino expressly stated is the basis for the research.
- Re-calculating statistical results with the unchanged values lowered significance of many entries and flipped the trend for the RAT_perf final score.
- Both the earlier version and the latest version of the data available for review were created in 2012 by Dr. Gino, and last saved by Dr. Gino, according to their Excel properties.

ANALYSIS AND OBSERVATIONS

Data Analysis.

The (a) data that was the basis for the published research, as indicated by Dr. Gino ("Publication Data"), was compared to (b) earlier versions of the data ("Earlier Data").

Observations.

Number of participants and demographics

According to the 2014 PS paper, p.977:

> *"Participants. One hundred seventy-eight individuals recruited on MTurk (47% male, 53% female; mean age = 28.59, SD = 7.72) participated in the study for $1 and the opportunity to earn a $1 bonus."*

When analyzing the data provided, 3.4% of the participants did not fill the 'age' question, and 2, Cum_IDs 185 and 192, did not complete the questionnaire and did not respond to the questions relative to pos_affect and neg_affect.. Hence, the results obtained for pos_affect and neg_affect are not based on 178 participants.

Discrepancies between conditions tested

The main two comparison groups in the statistical evaluations presented in the paper are: participants who cheated (cheaters) and participants who did not cheat (noncheaters).

When analyzing the *Earlier Data,* there is a column: "reported_guessed_correctly" that may be the source of the 'cheat' column, utilized to define which conditions had cheated [condition 1, cheat=1] or had not cheated, [condition 0, cheat=0] which the authors used for further analysis. There appear to be 12 entries whose values in 'reported_guessed_correctly' did not match the values in the 'cheat' column, and those values appear grey highlighted in the original document. See **Table 1** for a summary example.

Table 1. Screenshot of *Earlier Data* showing a portion of the columns of interest. (Black boxes added for emphasis of cells of interest.)

	Earlier Data		*Publication Data*
Cum_ID	reported_guessed_correctly	cheat [DO]	reported_guessed_correctly
99	1	1	1
36	1	1	1
97	0	1	1
153	1	1	1
8	1	1	1
171	1	1	1
136	0	1	1
191	0	1	1
105	1	1	1
32	1	1	1
51	1	1	1
187	0	1	1
172	0	1	1
9	1	1	1
157	1	1	1

159	1	1	1
180	1	1	1
14	0	1	1
146	0	1	1
20	1	1	1
156	1	1	1
149	1	1	1
31	0	1	1
169	1	1	1
63	1	1	1
102	0	1	1
129	1	1	1
11	0	1	1
42	1	1	1
138	1	1	1
77	1	1	1
170	1	1	1
22	0	1	1
74	1	1	1
84	1	1	1
128	1	1	1
106	1	1	1
35	0	1	1
50	1	1	1
38	1	1	1

Furthermore, the 0s for the IDs with cheat highlighted in grey in the *Earlier Data* appear all to be 1s in the newer version of the document, the *Publication Data* (see above **Table 1**).

In summary, it appears that within the Earlier Data, almost 30% of the reported cheating data [**reported_guessed_correctly**] appeared to be altered from condition 0 (=no cheat) to condition 1 (=cheat) in a separate column [**cheat**]. These specific alterations were apparently color-coded grey by an unknown individual handling the data. Furthermore, these altered data appeared to translate into the Publication Data in the reported cheating data [**reported_guessed_correctly**] The *Earlier Data* had only 31 observations for condition 1 (=cheat), while the *Publication Data* had 43 (27.9% more).

Discrepancies between RAT averages and manually entered scores:

When analyzing the *Earlier Data,* there is a column: "RAT_perf" that contains calculations. However, within this column a subset of data, 4 specific entries appear to have values manually entered. Similarly as above, for the conditions tested, these data have a grey background introduced by an unknown individual in the *Earlier Data*.

When these apparently manually inserted values are compared to calculated values using the formula from the category [=SUM(CVxxx:DLxxx)] and the available values for the calculation in the data sheet [for example, =SUM(CV170:DL170)] the apparent manual entries do not align with the calculated entries. See **Table 2** GREY vs RED cells for details.

Table 2. From *Earlier Data* (data and calculations) outlining a subsection of RAT_perf data.

CumID	**RAT_perf**	**RAT_perf** In 'Show Formulas' mode	**Calculated Values**
42	11	=SUM(CV168:DL168)	11
138	9	=SUM(CV169:DL169)	9
77	14	14	0
170	14	=SUM(CV171:DL171)	14
22	15	15	2
74	14	=SUM(CV173:DL173)	14
84	14	14	4
128	2	=SUM(CV175:DL175)	2
106	8	=SUM(CV176:DL176)	8
35	10	=SUM(CV177:DL177)	10
50	6	=SUM(CV178:DL178)	6
38	13	13	3

A close look at the Publication Data shows apparent modifications of the single RAT values that are then added to compute RAT_perf (see Table 3 left hand side).

Table 3. Example of data modified between *Earlier Data* and Publication Data

Cum ID	RAT																	SUM of single RAT values from Earlier Data	SUM of modifications	RAT values + modifications	RAT_perf reported
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17				
77	1	1	1	0	1	1	0	1	0	1	1	1	1	1	1	1	1	0	14	14	14
22	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	0	2	13	15	15
84	1	1	1	0	0	1	0	0	1	0	1	1	0	0	1	1	1	4	10	14	14
38	0	0	0	0	1	1	1	0	1	0	1	1	1	0	1	1	1	3	10	13	13

The sum of those modifications, when added to the RAT_perf calculated based on values from Earlier Data, results in the modified final RAT_perf values that were manually entered.

Of note, another version of *Earlier Data* shows no manually entered values (i.e., only calculated values) based on the single RAT entries. However, such calculations, given the editing of the single values themselves, ended up to sum to the manually entered value in the RAT_perf column. The modified entries, all part of condition 1 (= cheat), accounted for almost 10% of the condition 1 entries.

Therefore, it appears that the single values in the *Publication Data* were modified to SUM to the value reported as RAT_perf.

Statistical evaluation of the impact of the data modifications above

The statistical results described in the paper were calculated for the datasets considered, and the impact of the modifications above were also included.

Once the '**reported_guessed_correctly**' column from the *Earlier Data* was utilized to determine the data for the condition 1 [=cheat] the data for 'Caring about rules' and 'flexibility' became less significant than reported, and the 'RAT_perf' nonsignificant. 'RAT_perf' was also nonsignificant when re-calculating and applying consistent calculations across the dataset (see Table 4 row *Earlier Data – 'reported guessed correctly' as 'cheated' column and re-calculated RAT*).

An additional relevant effect the modifications had was on the averages and standard deviations for the two groups, for all the quantities estimated and, more so, for the RAT perf for which better RAT averages were measured for non-cheaters, but reported for cheaters, see **Table 5**.

Table 4. Summary of statistical outcomes, p value deviations

	Caring about rules	fluency	flexibility	RAT perf
2014 PS Paper	<.001	<.001	<.001	0.012
Earlier Data – 'reported guessed correctly' as 'cheated' column	0.002	3.58E-04	0.001	0.803
Earlier Data – 'reported guessed correctly' as 'cheated' column and re-calculated RAT				0.260

Table 5. Summary of statistical outcomes, means and standard deviations

	number of RAT items solved				caring about rules			
	Cheated		non cheaters		cheated		non-cheaters	
	M	SD	M	SD	M	SD	M	SD
2014 PS Paper	9.47 N=43	4.4	7.84 N=135	3.4	3.66 N=43	1.8	5.28 N=135	1.3
Earlier Data – 'reported guessed correctly' as 'cheated'	8.39 N=31	4.5	8.20 N=147	3.5	4.11 N=31	1.8	5.05 N=147	1.5
Earlier Data – 'reported guessed correctly' as 'cheated' and recalculated RAT	7.29 N=31	4.5	8.12 N=147	3.5				

Summary.

The analysis of the Earlier Data shows a series of data manipulations, a number of which were highlighted by an unknown individual indicating specific cells that have been modified. One example is the change of conditions for 12 entries where the 'reported_guessed_correctly' score appears to be "0" in the *Earlier Data* but become "1" highlighted in grey when apparently copied to the 'cheat' column used for the analysis. The same entries are then potentially modified all to be "1" for Publication Data, the analysis file identified by the respondent.
A second example is the manually entered values for RAT_perf in the *Earlier Data*, also highlighted in grey, in a column of calculations (Excel calculated SUMs of entered scores). In the Earlier Data, apparently RAT total scores were increased [changed to 1] to match the RAT_perf value that was manually entered (and eventually included in the Publication Data).

Notably, those modifications impact significance for 'Caring about rules' and 'flexibility', by lowering it, and the 'RAT_perf', which becomes non-significant. Additionally, the differences in averages for the two groups, for all the categories estimated, diminished while the trend of averages for the RAT_perf was inverted when re-calculating with the newly estimated data (see Table 5, Mean(cheaters, published)=9.47, Mean(noncheaters, published)=7.84 becomes Mean(cheaters, Earlier Data-reported_guessed_correctly)=7.29, Mean(noncheaters, Earlier Data - reported_guessed_correctly)=8.12).